

**RDAP 2018**

# Launching the Data Curation Network

**Lisa Johnston** University of Minnesota

**Jake Carlson** University of Michigan

**Cynthia Hudson--Vitale** Washington University

**Heidi Imker** University of Illinois

**Wendy Kozlowski** Cornell University

**Robert Olendorf** Penn State University

**Claire Stewart** University of Minnesota

**Mara Blake** Johns Hopkins University

**Joel Herndon** Duke University

**Elizabeth Hull** Dryad Data Repository

**Timothy M. McGeary** Duke University

03-22-2018

Data Curation Network

## Challenges



Created by Luis Prado  
from Noun Project

- Researchers are faced with a growing number of requirements (and incentives) to ethically share their research data.
- Well curated data are more valuable.
- The skills and expertise required to curate data cannot be fully automated nor reasonably be provided by a few experts siloed at single institutions.

The Data Curation Network (DCN)  
addresses these challenges by  
**collaboratively sharing data curation staff**  
across a network of partner institutions and data  
repositories.

# Planning Phase (2016-2017)



Alfred P. Sloan  
FOUNDATION

1. Compared local policy, technologies, and workflows across the 6 planning phase institutions
2. Held 6 focus groups with researchers on what data curation activities were important
3. Ran controlled pilots of data curation workflows with 17 data curators to ID issues
4. Surveyed the 124 ARL institutions to gauge support for data curation services
5. Researched cost recovery models for sustainable data curation and repository services
6. **Held information exchanges** with leaders of successful collaboration projects
7. Analyzed 1 year of data types, disciplines, frequency, and curation levels (needed vs taken)

# Planning Phase (2016-2017)



Alfred P. Sloan  
FOUNDATION

## Baseline Assessment

	Pre-ingest Curation?		Mediated vs Self-deposit?		Post-ingest curation			
Workflow Steps by Institution	Consult only	Staging Area for deposit	Mediated deposit	Self-deposit	As needed	Review metadata only	Review files and metadata	Add DOI
Minnesota	X			X			X	X
Cornell	X		X*	X			X	X*
Illinois	X			X			X*	X
Michigan	X			X			X*	X*
Penn State	X			X				
Wash U	X		X	X			X	X

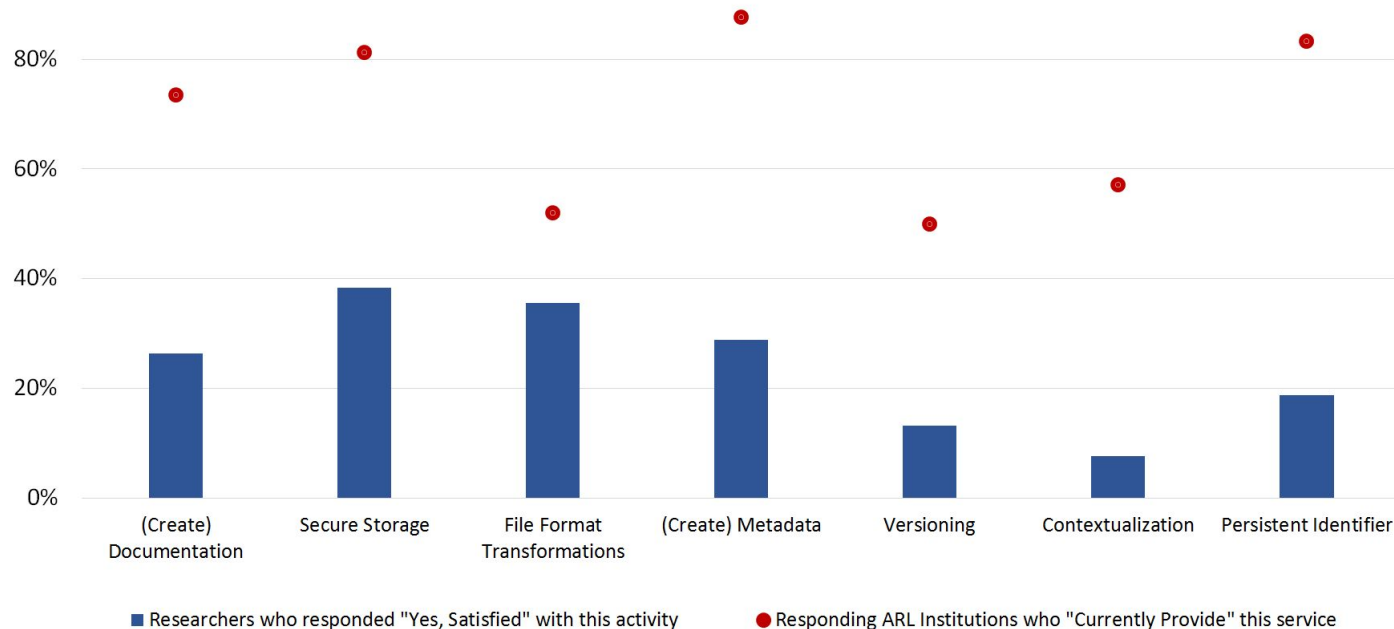
"Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." Journal of eScience Librarianship 6(1): e1102. <https://doi.org/10.7191/jeslib.2017.1102>.

# Planning Phase (2016-2017)



Alfred P. Sloan  
FOUNDATION

## Researcher Satisfaction and ARL Institution Curation Activities

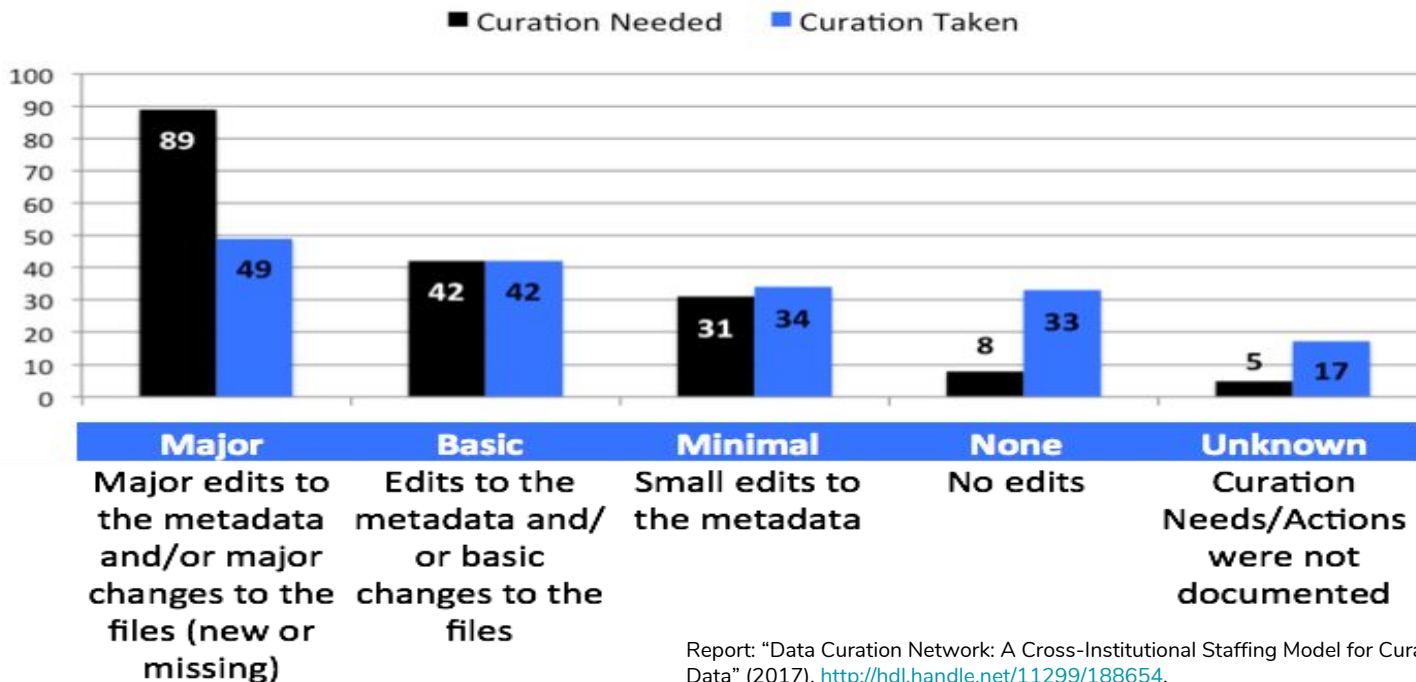


# Planning Phase (2016-2017)



Alfred P. Sloan  
FOUNDATION

## One Year Metrics Tracking (n=175)



Report: "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data" (2017), <http://hdl.handle.net/11299/188654>.

**The Data Curation Network (DCN) 3-year implementation phase is set to launch May 2018.**



Alfred P. Sloan  
FOUNDATION



# DCN Pilot Implementation

## 9 Institutions

- 8 Academic Libraries
- 1 General Data Repository

## 19 Data Curators

## 1 Project

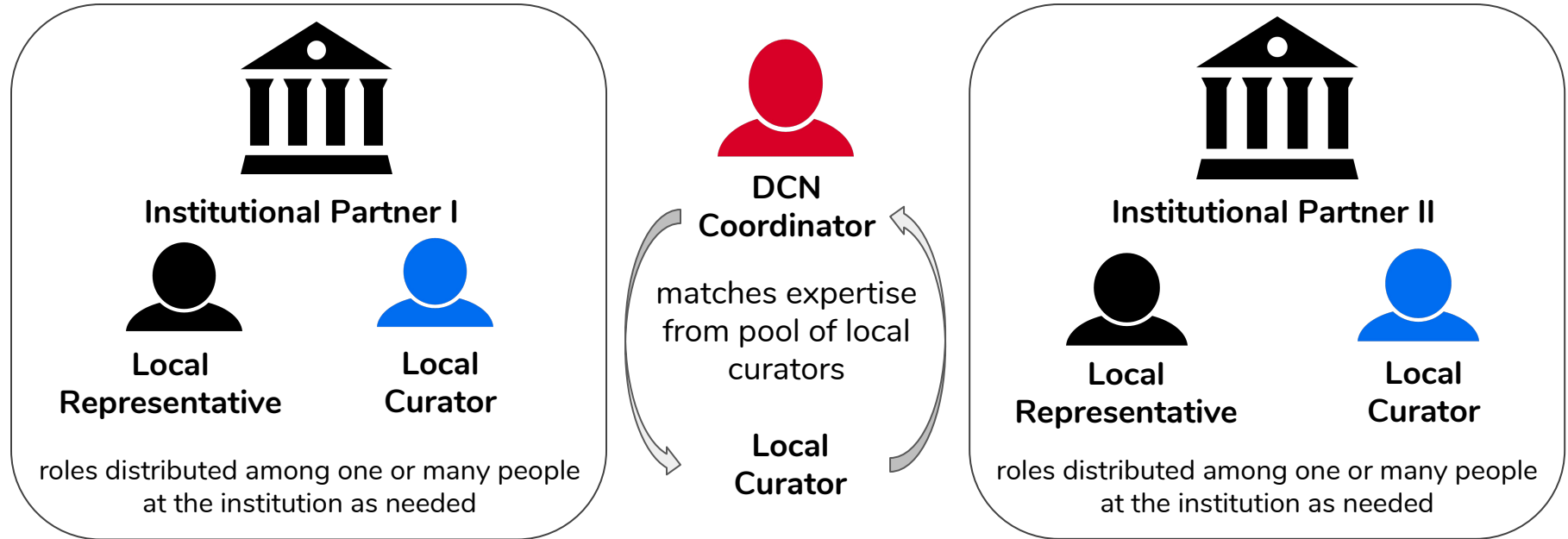
Coordinator (new hire)

1 Program Director (PI)



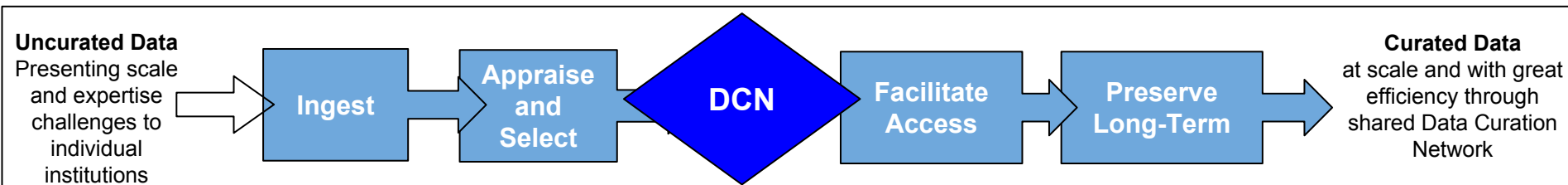
# DCN Alliance Model

Institutional Partners contribute staff and fund central coordinator.



Users gain access to the network via partnership or fee-for-service basis.

# DCN Workflow

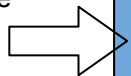


- Researchers deposit their data as they normally would
- DCN functions as a microservice layer (the “human layer in your repository stack”)
- Local institution maintain full responsibility for all technical functionality (eg. storage) and authority for local decision-making (what to ingest, how long to retain, etc.)
- Seamlessly integrates into all repository systems (Samvera, Fedora, DSpace, etc.)

# DCN Workflow

## Uncurated Data

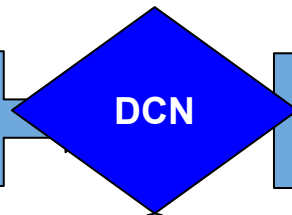
Presenting scale and expertise challenges to individual institutions



Ingest



Appraise and Select



DCN

Facilitate Access



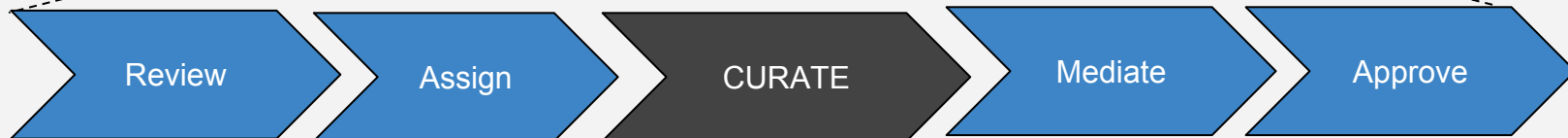
Preserve Long-Term



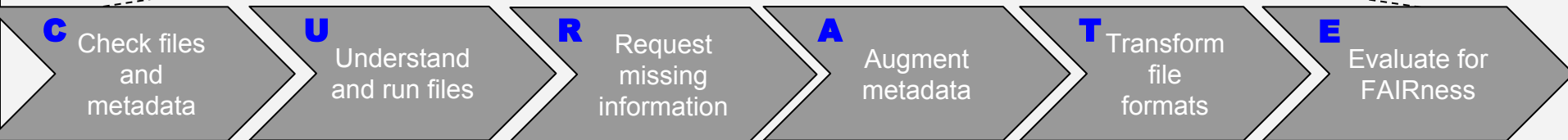
**Curated Data**  
at scale and with great efficiency through shared Data Curation Network

## Data Curation Network

### DCN Coordinator Workflow



### DCN Curator Workflow



# DCN Workflow

## CHECK Step

CURATE Action	Curator Checklist
<p><b>Check</b> data files and read documentation</p> <ul style="list-style-type: none"><li>Review the content of the data files (e.g., open and run the files or code).</li><li>Verify all metadata provided by the author and review the available documentation.</li></ul>	<ul style="list-style-type: none"><li><input type="checkbox"/> Files open as expected<ul style="list-style-type: none"><li><input type="checkbox"/> Issues _____</li></ul></li><li><input type="checkbox"/> Code runs as expected<ul style="list-style-type: none"><li><input type="checkbox"/> Produces minor errors</li><li><input type="checkbox"/> Does not run and/or produces many errors</li></ul></li><li><input type="checkbox"/> Metadata quality is rich, accurate, and complete<ul style="list-style-type: none"><li><input type="checkbox"/> Metadata has issues _____</li></ul></li><li><input type="checkbox"/> Documentation Type (circle) Readme / Codebook / Data Dictionary / Other: _____<ul style="list-style-type: none"><li><input type="checkbox"/> Missing/None</li><li><input type="checkbox"/> Needs work</li></ul></li></ul>

# DCN Pilot Implementation (2018-2020)

## Assessment Plan (two-prong)

*Is a networked approach to curating research data more efficient?*

- Number of datasets
- Frequency (high-volume time periods, etc.)
- Variety (data file formats; range of disciplines)
- Efficiency (time, costs)

*Are curated data are more valuable?*

- Track reuse indicators (download counts, citations, alt-metrics)
- Implement a DCN registry
- Apply badges and metadata to signal that data sets curated by the DCN are FAIR.

# 6-year Roadmap toward Sustainability

*Transition from planning phase to sustaining phase*

	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Support	Sloan Grant	Grant Funded (Y1-Y2) transition to partnership model (Y3)			Curation-as-service (Y4-6)		
Timing	2016-17	2017-19		2020-22		2022-2023	
Phase	Planning	Implementation		Transition		Sustaining	
Partners	6 academic institutions	8 academic institutions and a disciplinary partner			Recruit new partners as use and demand dictate		

**Mission: With a proven and appealing value-proposition, the Data Curation Network will expand into a sustainable entity that grows beyond our initial partner institutions.**

Thanks!

<https://sites.google.com/site/DataCurationNetwork>

Twitter #DataCurationNetwork